

学校编码: 10384

分类号_____密级_____

学号: 31520081153341

UDC _____

厦门大学

硕士学位论文

基于局部时空特征的人体行为切分与识别
方法若干关键技术研究

Research on Key Techniques in Human Action Segmentation
and Recognition base on Local Spatial-Temporal Features

刘 姝

指导教师姓名: 李 绍 滋 教授

专 业 名 称: 计算机应用技术

论文提交日期: 2011 年 月

论文答辩时间: 2011 年 月

学位授予日期: 2011 年 月

答辩委员会主席: _____

评 阅 人: _____

2011 年 月

厦门大学博硕士论文摘要库

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。
(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学博硕士论文摘要库

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

()1.经厦门大学保密委员会审查核定的保密学位论文，于
年 月 日解密，解密后适用上述授权。

(☒)2.不保密，适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人(签名)：

年 月 日

厦门大学博硕士论文摘要库

摘要

人体行为识别是计算机视觉领域的研究热点和难点，在智能监控、高级人机交互、体育运动分析和基于语义的视频检索等方面有着重要的意义和广阔的应用前景。计算机理解人体动作和行为将成为未来高层计算机视觉技术发展的重要方向。目前，人体行为识别的研究尚处于初级阶段，尽管在单一背景的简单动作识别上取得了一定的进展，但是由于人体行为的复杂性、场景的多样性等问题，真实场景下的行为识别仍然面临着许多困难。针对上述情况，本文围绕基于视频的人体行为识别展开研究，进行了以下三个方面的工作：

1. 归纳总结现有方法。查阅大量文献，对国内外计算机视觉领域的人体行为识别方法进行全面的归纳和总结，详细介绍了常用方法的基本思想和原理，并分析比较了各方法的优缺点。
2. 提出了一种基于熵的特征筛选方法。针对局部时空特征提取存在噪声和干扰等问题，本文通过计算特征点之间的熵，对检测到的局部动作特征进行筛选，剔除不良码本项，同时结合局部特征描述算子大多使用直方图进行特征表示的特性，使用改进的直方图相交核作为特征相似度的衡量方法。在 KTH 数据集上的实验结果表明该方法增强了特征码本的判别能力，提高了识别效果。
3. 提出了基于条件随机场的连续行为切分识别模型。针对复杂行为切分困难和识别率不佳的问题，本文借鉴中文词性标注中分词与标注同时进行的方式，将条件随机场模型与支持向量机相结合，建立连续行为的切分识别模型。通过对输入的图像序列进行基于光流的关键帧检测，将图像序列切分为动作基元片段，并对其进行局部特征检测和描述，构建运动特征码本。并结合支持向量机对动作基元的码本向量进行分类，从而使用条件随机场模型对得到的动作基元类别序列进行切分和识别，实现了模拟真实场景中，在不同的背景、尺度、视角和光照等条件下的连续行为的切分和识别。

总之，本文在总结现有人体行为识别方法的基础上，进行了简单行为识别和连续行为切分与识别等方面的研究工作，提出了基于熵和直方图相交核的动作特征筛选方法和基于条件随机场的连续行为切分识别模型，实验结果验证了本文所

提出方法及模型的有效性。

关键词：人体行为识别、特征筛选、连续行为切分与识别、关键帧提取、动作基元

厦门大学博硕士论文摘要库

Abstract

As a hot topic in the field of computer vision, human behavior recognition could be widely used in many practical applications range from intelligent video surveillance, advanced man-machine interface, sports motion analysis, and video retrieval based on semantic. Computer understanding of human action and behavior will be an important direction of high-level computer vision technology development in the future. Currently, research of human behavior recognition is still in its infancy. Although having a good performance in simple background and simple action, human behavior recognition in real scenes is still faced with great challenges because of the complex human behaviors and variable scenes. In response to these problems, in this paper we focus on the research of the following three aspects on human behavior recognition in videos:

1. Summarize existing human behavior recognition methods, and classify the methods of feature extraction and action recognition. This paper presents the basic idea of common algorithms in details and compares their advantages and disadvantages.
2. Propose a feature selection methods based on entropy for issues such as noise and other disturbances in local spatial-temporal features extraction. In order to remove bad visual words from feature codebook, it computes the entropy of each two features combined with histogram intersection kernel to measure feature similarity. Experiment in KTH database shows that our method could promote discriminant ability of visual codebook, and enhance the effect of action recognition.
3. Propose a CRFs-based model of segmentation and recognition for continuous behavior to solve problems of difficult segmentation and poor recognition in complex behavior. We are inspired by Chinese POS tagging, and combine CRFs with SVM to build up continuous behavior segmentation and recognition model. Our method segment video into primitive fragments of action by detecting optical flow and finding key frames from image sequence,

extract and describe local spatial-temporal feature in action primitive fragments to build visual feature codebook and get label of fragments by support vector machine. Using conditional random fields to segment and recognize continuous behavior from label sequence of fragments under different conditions including background, scale, perspective and illumination in the simulation scenes.

In summary, this paper research simple action recognition and continuous action segmentation and recognition based on existing methods of human action recognition. It proposes an action feature selection based on entropy and histogram intersection kernel, and a CRFs-based continuous behavior segmentation and recognition model. Experiment results show that our methods and model are effective.

Key Words: Human Behavior Recognition; Feature Selection; Continuous Behavior Segmentation and Recognition; Key-frame Extraction; Action Primitive Fragment;

目 录

第一章 绪论	1
1.1 选题背景及研究意义	1
1.2 国内外研究现状及存在的问题	2
1.2.1 国内外研究现状.....	2
1.2.2 存在的问题.....	4
1.3 本文主要研究工作	5
1.4 本文组织结构	7
第二章 人体行为识别方法概述	9
2.1 特征提取	10
2.1.1 背景差分法.....	10
2.1.2 时域差分法.....	11
2.1.3 光流法.....	12
2.2 特征表示	12
2.3 行为识别	15
2.3.1 基于模板匹配的方法.....	16
2.3.2 基于状态转移图模型方法.....	17
2.3.3 基于语法的方法.....	20
第三章 基于局部时空特征的动作识别	23
3.1 局部时空特征	24
3.1.1 局部时空特征检测.....	24
3.1.2 局部特征描述算子.....	29
3.2 基于熵的特征筛选	34
3.2.1 直方图相交核.....	37
3.3 分类器	38
3.4 实验结果与分析	38
3.5 本章小结	40

第四章 基于条件随机场的连续行为切分与识别.....	41
4.1 基于光流的关键帧检测.....	42
4.2 动作基元识别	44
4.3 构建模型	45
4.3.1 模型定义.....	47
4.4 实验数据集	48
4.5 实验结果与分析	49
4.6 本章小结	52
第五章 总结与展望	53
5.1 回顾与总结	53
5.2 未来研究工作展望	54
参考文献.....	57
致 谢.....	65
附录 攻读硕士学位期间发表的论文	67

Table of Contents

Chapter1 Introduction.....	1
1.1 Background and Significant.....	1
1.2 Present Research and Problems	2
1.2.1 State-Of-Art	2
1.2.2 Problems	4
1.3 Main Research Contents	5
1.4 Outline.....	7
Chapter2 Human Behavior Recognition Methods	9
2.1 Feature Extraction	10
2.1.1 Background Difference	10
2.1.2 Time Domain Difference	11
2.1.3 Optical Flow.....	12
2.2 Feature Presentation.....	12
2.3 Behavior Recognition.....	15
2.3.1 Template-based Methods	16
2.3.2 State-Transition Graph Methods	17
2.3.3 Syntax-based Methods	20
Chapter3 Action Recognition base on Local Spatial-Temporal Feature	23
3.1 Local Spatial-Temporal Feature.....	24
3.1.1 Local Spatial-Temporal Feature Extraction	24
3.1.2 Local Spatial-Temporal Feature Description	29
3.2 Feature Selection base on Entropy	34
3.2.1 Histogram Intersection Kernel	37
3.3 Classifier	38
3.4 Experiment Result and Analysis.....	38

3.5 Conclusion	40
Chapter4 Continuous Behavior Segmentation and Recognition base on Condition Random Fields	41
4.1 Keyframe Detection base on Optical Flow	42
4.2 Action Element Recogniton.....	44
4.3 Build Model	45
4.3.1 Model Definition.....	47
4.4 Experiment Database.....	47
4.5 Experiment Result	48
4.6 Conclusion	51
Chapter5 Conclusion and Prospect.....	53
5.1 Conclusion	53
5.2 Prospection of the Future work	54
References	57
Acknowledgements	65
Appendix Published Papers	67

第一章 绪论

1.1 选题背景及研究意义

视觉是人类认识世界、获取信息的主要来源，是人类交流中信息传递的重要方式。如何让计算机获得类似于人类的视觉感知能力，在图像及图像描述之间建立映射关系，从而使计算机能够通过数字图像处理和分析来理解视频画面中的内容，并采取相应的行动，是计算机视觉中一个极富挑战性的研究领域。

随着相机、摄像机、监控摄像头等数码产品大量进入生活，以及影视、互联网等媒体的迅速发展，视频已经成为信息传播的主要载体，人们对于自动识别和理解视频信息的需求也不断高涨。如何有效地处理并理解这些数据，最大限度的利用视频中的信息，为生产生活提供便利，是计算机视觉领域的研究目标。其中，人体行为识别是近年来备受关注的前沿方向和研究热点，它是通过对视频中图像序列的静态和动态信息加以分析，从中检测和识别人的动作。人体行为识别在智能监控、高级人机交互、体育运动分析和基于语义的视频检索等方面有着重要的意义和广阔的应用前景。

1. 智能视频监控

高级人机交互即在智能化的环境中，计算机能够通过不通过键盘和鼠标，与人进行自如的交流。这要求计算机能够理解人的自然语言和手势、表情、头部动作等身体语言，达到人与人之间交流的效果，是人机交互的最终目标，也是人类多年以来的梦想。早期被广泛应用于人机交互的语音识别，由于容易受到噪声干扰和距离限制，使得人机交互效果受限，而人的身体语言则不受此限制，它对于人机沟通和场景理解有着重要作用。因此，对人体动作的准确分析能够使计算机更好的理解人的意图，实现高度智能化。

2. 高级人机交互

高级人机交互即在智能化的环境中，计算机能够通过不通过键盘和鼠标，与人进行自如的交流。这要求计算机能够理解人的自然语言和手势、表情、头部动作等身体语言，达到人与人之间交流的效果，是人机交互的最终目标，也是人类多年以来的梦想。早期被广泛应用于人机交互的语音识别，由于容易受到噪声干扰和

距离限制,使得人机交互效果受限,而人的身体语言则不受此限制,它对于人机沟通和场景理解有着重要作用。因此,对人体动作的准确分析能够使计算机更好的理解人的意图,实现高度智能化。

3. 体育运动分析

行为识别的另一应用是体育运动分析。在体育运动中,体育运动分析可用于创建个性化训练系统,对体育运动视频进行自动分析、评判,提供科学直观的分析数据,加速数字化体育运动训练进程。同时,它也可以为体育比赛的转播提供比赛评注、赛况自动降解等实时信息。

4. 基于语义的视频检索

随着互联网的发展,视频已经成为人们生活中不可缺少的部分。在海量视频数据库中查找具有某种语义信息的视频信息变得非常困难。传统的检索方法是通过给视频添加文字或数字标签,然后在索引时,对标签进行检索。但是,视频内丰富的信息很难用标签完全表达,且标签容易受标注者的主观因素影响,同时,大量的待标注视频需要耗费巨大的人力资源。因此,基于语义的视频检索需要人体行为识别对视频内容的自动标注,实现对视频数据库的高效管理和查询,提高信息的使用效率。

正是由于人体行为识别具有广泛的应用前景,使得这一研究领域受到广泛关注,目前国内外还处于研究探索阶段,实际应用相对较少,其潜在的巨大经济价值和社会价值尚未得到开发。大部分人体识别的研究将场景设定在一个受控环境下,如静态背景、无遮挡等,对简单场景下的人体行为识别已经达到 90%以上的正确率。但是在真实场景中,由于受到光照变化、阴影、视角不同、尺度变换和复杂背景等诸多因素的影响,人体行为识别仍然存在许多难题有待解决。因此,本研究课题对该领域的关键技术问题进行探索和研究具有重要的理论和现实意义。

1.2 国内外研究现状及存在的问题

1.2.1 国内外研究现状

最早利用序列图像对运动人体进行跟踪与分析的研究可以追溯到 1973 年,

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库